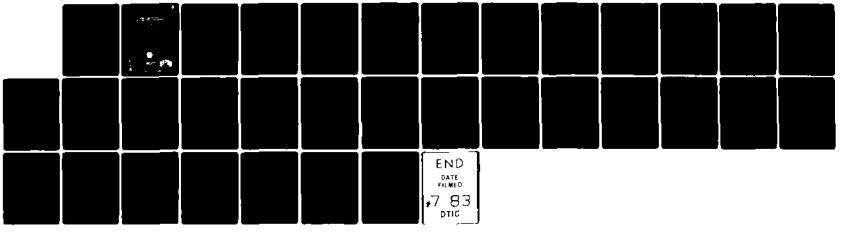
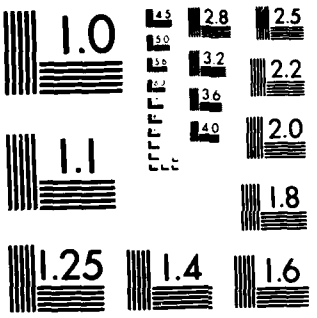
AD-A129 466 DESIGN OF AUDIO CIRCUITS FOR INPUT-OUTPUT OF DIGITAL VOICE PROCESSORS(U) NAVAL RESEARCH LAB WASHINGTON DC D C COULTER 21 JUN 83 NRL-MR-5100 1/1

UNCLASSIFIED F/G 17/2 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NRL Memorandum Report 5100	2. GOVT ACCESSION NO. AD-A129466	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DESIGN OF AUDIO CIRCUITS FOR INPUT-OUTPUT OF DIGITAL VOICE PROCESSORS		5. TYPE OF REPORT & PERIOD COVERED Final report on a continuing NRL problem. 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) D. C. Coulter		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, DC 20375		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 28010N; X0919-CC; 75-0126-0-3
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Electronic Systems Command Washington, DC 20360		12. REPORT DATE June 21, 1983 13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Audio; Voice processor; Input circuits; Output circuits; Voice circuits; ANDVT; Automatic Gain Control (AGC); Sidetone circuits		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Good design of input-output audio circuits for mostly digital voice transmission systems, such as the Advanced Narrowband Digital Voice Terminal (ANDVT) may be neglected due to commonly held misconceptions about voice circuits which do not apply here and by a lack of sufficient audio design training of digital engineers. This report details pitfalls and reference sources for this specific design area and also details considerations for design of Automatic Gain Control and Sidetone Circuits.		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CONTENTS

Design of Audio Circuits for Input-Output of Digital Voice Processors

1.0 Overview

It is commonly assumed that audio circuits used for speech reproduction can be designed without regard to fidelity as long as nominal transmission is provided in the 300 to 3000 Hz range. This assumption is justified by the notion that speech intelligibility rather than quality is important (and extensive tests have demonstrated that excellent intelligibility is preserved by transmitting this band) and the idea that non-linear amplitude distortion is less damaging to speech quality than it is to music quality (a fact that can readily be demonstrated by attempting to play music through a carbon transmitter of a normal telephone).

While the above observations may be correct for a telephone system, they are not necessarily applicable for the input or output portions of a voice system which analyzes and synthesizes speech as a means of bit rate reduction (such as VOCODERS and Linear Predictive Coders (LPC) devices). The reason is, that however sophisticated such devices may be, they do not employ the same identical mechanisms as the human ear, and even though the ear is the final output device, it is "insulated" from the input

Manuscript approved March 29, 1983.

by the analysis-synthesis system. An example of this was the requirement of early (analog) vocoder designs for flat microphone/input circuit response down to 60 Hz to allow pitch tracking of fundamental pitch. In contrast, experiments have demonstrated the ear's ability to perceive low frequency repetition rates in signals high-passed as high as 2000 Hz [1]. More modern processors have ameliorated the need for fundamental pitch somewhat, but other types of special requirements still remain. Since it is well within present "state of the art" to provide good audio performance, it seems desirable to do so, provided a reasonable assessment of what "good" is can be provided.

The purpose of this report is to delineate such requirements as they are known or can reasonably be inferred and to describe the rationale used to obtain them. In order to do this, certain concepts will be introduced which, while they are not completely absent from the literature, are not contained in the normal electronic engineering curriculum. Whenever possible, quantitative standards will be suggested as the basis for ongoing development of these systems.

2.0 Subjective Considerations

How can one judge the qualities of a voice signal reproduced in various common ways (from a tape recorder, through a loudspeaker, from a telephone, face to face in a noisy environment, possibly in partial or total darkness)? This task, which first appears to be easy, is later found to be fraught with difficulty.

Nature has endowed man with many aids to voice communication, and they may all be available to a greater or lesser extent in a given communication situation. These include redundancies in language (which may be increased purposely in certain structured communications environments), the ability to

lip read * and the ability to sort out a given conversation in a noisy background by making use of two-eared correlation [2]. Nature has also made it difficult for us to evaluate the help we are getting except to the obvious extent to which the speech becomes difficult to understand (or downright unintelligible); otherwise, we tend to be blissfully unaware of all the processing being applied to a given situation. Even with a system which produces sometimes severe amplitude and frequency distortion, such as the telephone, can be adapted to, and so it is not uncommon to speak of a particular person's "telephone voice" as a learned response to these distortions on that particular voice.

Why is it important to deal with these commonly accepted aspects of vocal communication in a report of this type? Because it is these effects which make it difficult to realize the degradations in common audio and reproduction systems, and hence to deal with them as they may affect voice analysis system performance.

On the assumption that the reader is still skeptical, another example is given below:

It is common to believe that the distortion of a good high fidelity tape recording of a voice is negligible. For example, on machines capable of instantaneous playback, an A-B comparison of the input signal before recording and the signal played back

---

* This is practiced consciously by the deaf or partially deaf and also unconsciously by most other people, but cannot be done in darkness.

(a few milliseconds delayed) will show little discernable difference between the two (except slightly more background hiss on the recorded version if played on a super-fidelity system). However, the following simple test will demonstrate the degradation which actually occurs. Simply record a single word and feed the output of the playback back to the record input at about the normal record level, and let the "captured" word re-record itself many times. The effect, at first sounding like an echo, will quickly degrade, in 5 or 10 re-recordings, into something which is totally distorted. And yet, this distortion is not detectable after a single, or perhaps 2 or 3 copies! It is obvious that degradation is occurring, it's just that its effects have to exceed a certain threshold to be noticed by the human ear.

The foregoing discussion illustrates that considerable distortion can occur to speech before being noticed *, and the reason that this fact is so important is that it leads to a relaxed idea of what is important for audio design for speech circuits, particularly if they involve digital analysis-synthesis processing. The logical consequence of this

---

* It has been determined that distortion in music is more perceptable than speech. This is probably due to the simpler harmonic structure of speech where all frequencies are related to the pitch or larynx vibration rate, so any cross products do not produce inharmonic sounds, but only disturb the relative harmonic amplitudes. The ear is relatively insensitive to this, but the same is not necessarily true of equipment. One additional reason for the ear's insensitivity is that it is still able to pick out formant peaks which carry intelligibility. But in this case, these peaks are further distorted due to LPC or vocoder transmission.

could be to design only the best possible system (total Hi-Fi concept). While this may work, it is probably not necessary in the majority of cases (some engineering tradeoffs may also be essential due to practical constraints) and it is certainly not the "smartest" way to go, however "safe" it may be. The remainder of this report is to delineate as much of the known relationship between audio design and system performance as possible. In addition, many common pitfalls will be discussed. This will be done with the goal of making specification choices on as rational a basis as possible and to provide room for improvement as further knowledge becomes available. Another goal is to instill proper respect for the problems involved, and to insure that proper emphasis is given in proposal preparation, funding, and development personnel assignment.

3.0 Development of Individual Circuit Function Criteria

3.1 Input Coupling

In most cases, the audio will be routed from a subscriber terminal (for example, located in a CIC) via cable to the voice processor, frequently travelling through a switchboard. These circuits are typically balanced 600 ohm audio lines (usually grounded center tap) although in some cases they may be unbalanced (one side grounded). Therefore, if the input to the voice processor is designed as <u>only</u> balanced or <u>only</u> unbalanced, there is a good chance it may be wrong for a given installation. This can be avoided by making the input changeable, but there is a good chance that this installation option will be improperly selected by the installer. When this occurs, hum pickup, low level, loss of low frequency, etc. results, but the system still partially works, resulting in a usable installation but with marginal performance. To avoid this possibility, many systems have a "floating" input provided by an input transformer. This

recommended procedure * works equally well for both balanced and unbalanced inputs, and is not too critical to proper impedance match. (see end of this section for impedance matching considerations.) The major pitfall is in specifying the design of the input transformer. This is especially true in the design of modern equipment since there is usually strong pressure to reduce the size of the transformer, and there are definite limits to how much reduction is possible without hurting performance. The situation is complicated by another factor: transformer manufacturers, seeking to make transformers more attractive for use in miniaturized circuits so as not to lose their markets, have overdone their size reduction in some cases and sacrificed performance to a drastic degree, and then listed their products for "voice applications" while carefully not stating the actual, sometimes shockingly poor specifications. As stated earlier, there is such tolerance by the ear of such distortions as to actually make these products useful, but in the applications discussed here, they may literally destroy the capability of a voice processor. For this reason, one should never rely on a transformer without specifications, or with published response but not the power level measured, or one merely "recommended for voice applications."

On the other hand, the requirements for voice, when precisely understood, are not as severe as those for, say, high fidelity music, and so there is some degree of miniaturization which is perfectly acceptable once the requirements are understood.

---

* One might also think of choosing a solid state device such as a differential input Op amp as an input device. It is the writer's opinion that this usage, however attractive from a size/weight standpoint, would be difficult to implement in a way that is sufficiently protected from transients and RF to not be a potential failure point in the system.

In the first place, the lowest frequency in voice is much higher than music. Where for the latter a low frequency limit of 20 Hz is commonly used (the lowest organ notes heard by the human ear as tones), the voice fundamental (pitch frequency) is rarely lower than 75 Hz and usually much higher (above 150 Hz for most females). In addition, although early vocoder designs utilized this fundamental component for tracking pitch, and hence required input circuits (and microphones) to respond to about 60 Hz, more modern circuits such as the AMDF concept * do not require the presence of the fundamental, but operate mainly in the 200 to 800 Hz range of the lowest ("first") speech formant. Two other facts are relevant here. The first is that typically, the fundamental frequency is about 6 dB lower than the second harmonic because of the assymetrical nature of the vocal pulse (shaped somewhat like the output of a 1/2 wave rectifier) and so the power is reduced for this component. The second is that the fundamental carries little information about the speech signal (the ear appears to listen for harmonic spacing to determine pitch, not fundamental, and the lowest voice resonances are about 275 Hz for the first formant of "ee" and about the same for the nasal resonance of "m" and "n" nasal sounds.) For this reason, it has long been a practice in voice communications systems to limit the low frequency response, usually to start rolling it off below about 300 Hz. While this reduces the naturalness of the voice, it avoids using transmitter power for the lower frequencies and may actually improve received signal to noise ratio by a few dB for this reason, and thus improve net performance. There is a danger of carrying this too far however, and in particular, too high

---

* AMDF = Average Magnitude Difference Function uses the speech signal delayed and subtracted from itself to locate pitch epochs and does not depend on the presence of the fundamental frequency. Other correlation-based pitch tracking systems are likewise insensitive to fundamental.

a cutoff will reduce the ability of the listener to discriminate nasal sounds, so it is suggested that cutoffs above 250 Hz be avoided for best intelligibility.

Returning to the input transformer design, there is a hazard in over-interpreting the above limits regarding low frequency inputs. This is because the fundamental and second harmonic will still be present in the input speech from many of the microphones in use, and hence, even though these frequencies are not important to voice processor performance, they may create distortion at higher frequences or even saturate the transformer core so greatly as to prevent transmission at higher frequencies. It should be pointed out that transformer size is determined by the lowest frequency to be transmitted at maximum level; for most speech systems this will probably be the second harmonic of the larynx frequency, (since the fundamental is typically 6 dB lower as stated earlier) which has a range of about 2x75 to 2x200 Hz (150 Hz to 400 Hz) for males and 2x140 to 2x300 Hz (280 to 600 Hz) for females. When the first formant (about 275 Hz for "ee" or "m" sounds) is located on top of the second harmonic of pitch, this represents the probable worst case leading to a lowest frequency requirement at 275 Hz and not 150 Hz. This is the criteria for transformer design used for the Advanced Narrowband Digital Voice Terminal (ANDVT) [3], where a criterion of +10 dBm at 275 Hz with less than 0.8% distortion (distortion - 42 dB below signal) was established for the input circuit, including transformer (the other parts of the circuit should be much better than this); this insures that the transformer will not saturate with any speech input and thus cause distortion products in any part of the audio range. While this would seem a relatively lax requirement (compared to "Hi Fi" transformers responding to 20 Hz) it does not result in as tiny a transformer as is typically sold for "voice transmission" even when best state-of-the-art transformer core materials and design techniques are used. A different and better "off-the-shelf" transformer for this type application is the "Telephone Interconnection Transformer" marketed by several manufacturers to be used with multi-tone

MODEMS feeding into telephone lines. Although not intended for voice, they are suitable for voice since MODEMS are designed to exploit maximally the existing voice telephone lines, as well as to transmit many tone frequencies simultaneously all at about the same power, with low intermodulation distortion to avoid errors. Hence it is not too surprising that they are a bit larger and much more tightly specified than "voice transmission" transformers, and these units could be used, perhaps with some additional shielding, for input to voice processors. The typical specifications for these transformers are 275 to 3,500 Hz $\pm$ 0.5 dB response at +7 dBm with distortion a minumum of -46 dB below the signal. Some models which are specified to 4,000 Hz and +10 dBm would be even more satisfactory. The point here is that the necessity for carrying the lowest MODEM tone at full power, coupled with the necessity for flat reproduction of all the tones across the band produces a transformer which also satisfies the primary requirements for the input to a voice processor, and the availability of such a transformer proves the feasibility of meeting these specifications in a reasonably small package.

The following paragraphs address the subject of input impedances and matching. Impedance matching in the audio world is a much misunderstood concept, with the result that unnecessary things are done in its service which would not be perpetrated if the fundamentals were better understood. The most prevalent misconception is that it is necessary to match a short transmission line which carries only audio frequencies. Transmission line theory, which is of great value for electrically long lines (i.e., lines which are about 1/8 wavelength and longer), is of little value in audio usage except for telephone transmission of some distance. This is because the wavelength of the highest audio frequency of interest (about 4,000 Hz) is in the order of 46 miles, giving a length for 1/8 wavelength of about 6 miles. Thus for most audio lines of lengths of a few feet to a few hundred feet, termination of the

line in its characteristic impedance (determined by conductor size and spacing) has little usefulness unless the power is needed at the end. The electrical capacitance of the line may have the effect of rolling-off the high frequencies if the source or driving impedance is too high or the capacitance is excessive (as may be the case with some types of shielded cable). However, terminating such an electrically short line in a resistor is generally a waste of power and serves no useful purpose. Sometimes, persons become confused by the power matching concept, which says essentially that the maximum power is drawn from a source of given impedance when this impedance is matched by a like quantity. This idealistic concept is rarely met in practice, since most active drivers (such as vacuum tubes, transistors, Op amps) work most optimally with loads far different than their source impedance (generally many times greater) and their source impedance is kept low by unintentional (emitter or cathode coupling) or intentional (external loop) negative voltage feedback. In these cases, usually the most important consideration is whether the driver can develop enough current in both positive and negative directions to charge the capacitive load presented by the circuit at the highest frequency (generally the aforementioned shielded cable). The inability of circuits to handle such currents is quite common in practice and underlies the recently rediscovered distortion known as "slew rate" or transient intermodulating distortion [4].

Of course there are certain situations where impedance matching is important, even crucial. These are as follows:

(a) Passive filter inputs and outputs
(b) Matched attenuator inputs and outputs
(c) Audio transformer inputs and outputs
(d) Driving a loudspeaker or other device requiring power

Even for the above four cases, there are many caveats and peculiarities.

For passive filters, the sensitivity of their characteristics to source and load impedance varies markedly from one individual filter to another. For example, some filters must be driven from precisely the specified source impedance and loaded likewise (say ± 10%). If they are not, their published characteristics will not be realized and may be drastically altered, with actual voltage gain at some frequencies. For these units, it is frequently necessary to insert series resistance between the driving amplifier or test oscillator and the filter, since these devices although designed to be loaded by a certain impedance, e.g., 600 ohms, actually have a source impedance which is far lower than this, and to drive the filter with this lower impedance will drastically alter its response. Many filters also have outputs which are designed to be loaded by a certain impedance. For such filters, 6 dB voltage is typically lost due to "matching". In other cases, loading is unnecessary, and thus this loss is avoided.

In the case of matched "broadcast type" attenuators (usually "T" or "H" attenuators) it may again be necessary to increase the source impedance of the driver and match the output to preserve the accuracy of the attenuation reading.

In the case of audio transformers, it is normally customary to stay within a 2:1 ratio of the rated impedance to preserve the frequency response and minimize transformer wire losses (transformers are designed with winding inductance which determines the highest impedance useful for its lowest frequency and wire size which determines its losses at lowest impedance used). However, to achieve this, it is only necessary for the line impedance to be proper, not that it be matched with a like impedance.

In the case of amplifiers used to drive power absorbing devices such as loudspeakers, there are two impedances to be considered. The first is the desired load impedance for lowest distortion and maximum power transfer, typically, a compromise between these factors, for a given speaker impedance. The second is the apparent source impedance, typically much lower as provided by feedback and helpful in damping mechanical resonances in the loudspeaker.

Having explained the above points about short line termination and other aspects of impedance matching, it is now possible to explain the situation with audio inputs to equipment.

For many years designers of audio broadcast and professional equipment have not adhered to strict impedance matching except where absolutely necessary as in some passive filters and broadcast-type attenuators. Persons in the business of designing audio systems understand this and the principles involved are derived from basic engineering, but as far as the writer is aware, there are few texts where this information is available and thus it has become part of the exclusive knowledge of experienced audio engineers [5].

Let us now take the most important case of "not matching" an input. Why is it necessary to match a relatively short 600 ohm line input? The answer is, it is not (unless the line is 6 miles long or longer for voice frequencies). Thus for running signals within an equipment space, there is no 600 ohm resistor used to terminate the lines. This practice has several advantages:

(1) 6 dB of gain is picked up that would otherwise be lost. *

---

* A few years back it was common practice to sell a broadcast line amplifier with an actual 34 dB of gain between 600 ohm inputs and outputs as a "40 dB amplifier" (High impedance input)

(2) It is not necessary for the line amplifier to deliver *any* <u>*real*</u> *power*, even though a nominal power of 0 dBm (1 milliwatt) or +10 dBm or more would be represented by such a voltage level into a real 600 ohm load. This was advantageous to the designer and also saves power. As a result, few, if any, devices like tape recorders, amplifiers etc. that are rated with 600 ohm outputs at so many dBm will really <u>drive</u> 600 ohms to that power level--in fact most would be pitifully inadequate to do so. Thus it is difficult to find devices that will really drive a 600 ohm passive filter or attenuator at levels of 0 dBm or +10 dBm.* Also, many outputs, even if they will drive 600 ohms at these levels, will be severely distorted in doing so, having been intended for a load of 10,000 ohms or greater. In some cases there may also be output coupling capacitors intended for no less than 10,000 ohm loads; loading these circuits with 600 ohms severely reduces low frequency response. Also, a circuit having a 600 ohm value of capacitive reactance (capacitive loading of about 0.06 microfarads at 4,000 Hz) will probably cause severe distortion because of the source's inability to deliver enough charge or discharge current to the load. Fortunately, most integrated circuit operational amplifiers are specified in this regard. But driving even moderately long lengths of high capacity shielded cable at normal voice levels of 10 dBm or so is not a trivial proposition. Of course for longer line lengths, the capacitance is distributed rather than lumped and becomes part of the characteristic

---

* A further complication is that a sine wave power capability of +10 dBm is required to transmit a voice signal over a measured 0 dBm, circuit, due to the voice waveform peak power to RMS power ratio.

impedance, and the amount of capacitance for such a line of 600 ohm impedance is much lower per foot than for shielded cable (even RF type shielded cable which is designed to minimize capacitance typically has an impedance of about 50 ohms rather than 600 ohms, reflecting its higher capacitance per foot.)

(3) In spite of the above considerations, there are times when matching is necessary in order to preserve an existing interface. For example, if a new high impedance piece of equipment is to replace an existing equipment which does match the line, a net increase of line level may occur, depending on the line's source impedance. This must be evaluated for each case. It is, however, recommended not to match unless it is necessary, and to explain the situation in the specifications supplied. In most cases in recent equipment, source impedances will be low, due to feedback, and so the above precaution will not be applicable.

3.2 Input Amplifier Considerations

The most important initial question to answer is whether to use discrete components e.g. transistors, capacitors, etc. or to use IC's, e.g. op amps with feedback, as audio amplifiers. The recent trend would be to use op amps; the discrete option requires some defending. This is because size considerations favor IC'S, and after all, no one would think of building a discrete digital circuit anymore. Then why even consider discrete circuits for audio? There are two reasons why, and unless these considerations are thoroughly understood, the road to using IC'S in this particular application is literally paved with land mines. The author has experienced more than one IC input circuit for a digital voice processor which was not only so bad as to be unsuitable, but which was virtually unfixable

without total redesign. The unsuitability, moreover, was readily measurable in lowered intelligibility scores, as well as audible effects. The two pitfalls are: (1) slew rate distortion [4] and (2) distortion due to the high sensitivity of op amp inputs to high frequency pulse interferences from nearby digital circuits.

Slew rate distortion, in brief, is a recently re-discovered phonemenon, now associated with operational amplifiers with a lot of negative feedback *. It is so serious in high fidelity applications that a whole generation of defective recordings were produced and marketed before this problem, unmeasurable by normal means, was discovered in the newly developed professional recording equipment using IC op amps of the early 1970's. Its discovery has led to the introduction of slew rate specifications on op amps, and the introduction of many new specialized types of op amps for audio use. Older type numbers such as the 709, 741, and 301 are probably not suitable for frequencies above about 1 kHz, and should not be used. Special high performance type numbers such as NE 5534 and 33, HA 2625, LM 318 and AD 518 are probably not required for this application, although they would be indicated for "HI FI" (up to 20 kHz and above) applications. Therefore, if IC'S are used, medium performance units such as LM 310, MC 1456, NE 531 are recommended. As design with op amps for good quality audio results is anything but

* Much earlier observation of this same effect occurred when a cathode follower (vacuum tube) was used to drive a highly capacitive cable.

results is anything but simple, the reader is directed to an excellent text to be studied before the design is attempted. This 200 page book [6] is practically must reading for the engineer of such a circuit: however, thorough as it is, it does not cover the second point to be raised here, namely, pickup of high frequency energy from nearby digital circuits. The most likely reason for this omission is that most audio circuits are not on the same chassis with high speed digital circuits, and such proximity is probably unique to digital voice processors.

Since this is not covered elsewhere insofar as is known, an attempt will be made here to describe the problem. In fig. 1 is shown a simplified diagram of an operational amplifier as used in an audio application. Since the amplifier has very high gain, (which is useful in making an accurate integrator) the ratio of the two resisters $R2/R1$ essentially determines the circuit gain, in these applications in the order of from about 1 to 30 or so. The point here is, that any high frequency signal picked up as shown by the arrow, due to capacitive coupling (from a nearby digital switching field which may be fairly intense) will not be passed by the op amp because the frequency is outside its bandwidth. Since there is no feedback at these high frequencies either, the gain in the early stages of the op amp, before bandwith limitations take over, is likely to be very high. This means that very low amounts of high frequency pickup across a small value resistor may overload the op amp internally and these small voltages may be very hard to detect with an oscilloscope. Since op amps are not normally operated in such a

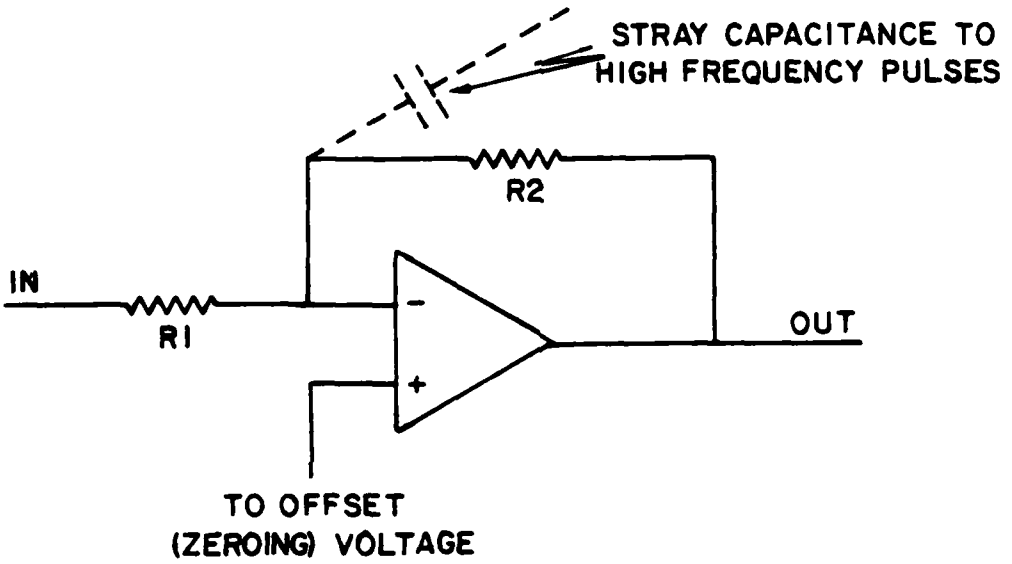


Fig. 1 — Op-Amp high frequency overload mechanism

field, it is a situation with which the manufacturer would not be normally concerned. What it means here is that the designer must use very short leads and shielding if he is to avoid such a potential problem. In the cases observed, the following effects of such pickup were observed:

(a) Excessive harmonic distortion, which did not gradually increase with audio level but changed erratically.

(b) Gross change in performance depending upon whether the audio card was on an extender or not.

(c) Unreliable frequency response and gain measurements.

It is thought that careful design with an awareness of this problem would avoid it and some more recent op amp audio front end designs have been successful. For example, careful by-passing might be used (where possible without upsetting stability). In the particular case described, the circuit was replaced with discrete components, which cured the problem. Using discrete components, the gain before feedback was modest, thus reducing the system vulnerability. Other fixes used for typical EMC problems, such as ferrite beads and shielding, are expensive when applied after construction; the best idea is to be aware of the potential problem and try to avoid it. It should also be pointed out that for the example cited the space required for discrete components was very little more than needed for IC'S because gains are moderate and circuits can thus be simple (i.e. 2 transistor amplifiers). Therefore it may be a practical alternative to use discretes for some systems in the audio circuits even though it would not be for any of the other parts of the system, and by so doing, avoid the problems just described.

Another design consideration for the audio input revolves around the large differences in peak-to-RMS ratio between sine waves, normally used for testing for frequency response and distortion, and speech waveforms, which consist of damped sine waves with peaks which are about 10dB higher than the peaks of equivalent-RMS-reading sine waves. The first time this problem appears is usually during the breadboard stage of development. Breadboards typically have an audio level or VU meter. (an AC meter, calibrated in dB with an arbitrary zero reading, usually +4dBm or 4dB above 1 mw across 600 ohms which equals about 1.2V RMS) * If the meter sensitivity is adjusted so that the A/D converter is fully loaded with sine wave input, i.e. the peaks of the sine wave drive the input to maximum range extremes without overdriving, then with speech input reading OVU on peaks, the A/D converter will be overdriven about 10dB (about 3:1 voltage). Few engineers who have not had audio design experience know this. Most audio engineers themselves learned it by experience rather

---

* This VU meter, used extensively by the broadcast industry was standardized because earlier "dB meters" were unsatisfactory due to having the "0" point or reference too low on the scale, and utilized too little damping of the movement, hence they had substantial error due to needle inertial overshoot on speech, and utilized too high a power level (6 mw instead of the present 1 mw) as a reference. Unfortunately, the new VU meter was designed by a committee, and it was later found that a 3,600 ohm series resistor was necessary when the meter was used across a 600 ohm audio line to prevent the distortion caused by the meter selenium bridge rectifier from exceeding 1%. This results in the reduced sensitivity of +4dBm rather than the intended 0dBm which is present when the meter is connected directly across a 600 ohm circuit.

than in engineering school, and they tend to specialize in designing audio equipment, not military voice processors. However, once realized, it is usually possible to increase the meter sensitivity by about 10dB and hence eliminate the overload to the A/D converter by decreasing the gain or talking level to achieve a readjusted "0" level on the VU meter. Having done this, another pitfall occurs. It now becomes necessary, in order to measure distortion, frequency response, etc. at realistic levels, to put sine waves in at levels that peg the VU meter by at least 7dB (the 3dB to the right of zero plus 7dB more equals 10dB). And because of this, input amplifier design of many systems will not even deliver full A/D input range without clipping; they were never tested to the same peak-to-peak voltage swing as the speech will subject them to, but to "0dB on the VU meter for sine wave input." To allow properly for testing, it may be necessary to provide a switch to reduce the meter sensitivity by 10dB or, to disconnect the VU meter. This maximum point should definitely be tested since most A/D converters have input requirements of $\pm$ 10V or $\pm$ 5V. These voltage swings may not be available from normal circuits due to insufficient supply voltage, particularly if a passive anti-aliasing filter, with a potential loss of 6dB or 2:1, is placed between the audio driver and the A/D converter. In that case, it will probably be necessary to place an additional amplifier after the passive filter to drive the A/D converter properly.

The last point is the question of audio response. As was stated earlier, the flatness of response in-band is more important than in conventional audio systems for speech because of the degradation occurring in the coding process. It should

not be a challenge to design an amplifier which is flat to within less than 0.5dB over the range of 200 to 4,000 hz. However, it must be considered in the design and measured to be certain that some inadvertent rolloff has not occurred, as for example by insufficient coupling capacitors on the low frequency extreme, or stabilizing feedback capacitors on the high frequency limit.

There is also another issue with regard to the input bandwidth: that is," Why is 4kHz used instead of the usual 3kHz speech band for the ANDVT voice processors?" Two things need to be pointed out. One is that telephone bandwidth of 3kHz is in itself a form of bandwidth compression, where the trade-off of less perfect intelligibility and quality has been found to be an acceptable one in exchange for the high cost of greater bandwidth. * After all, speech researchers since Fletcher have shown that at least 6kHz of bandwidth is necessary for "perfect" intelligibility (the ability to understand nonsense syllables passed through the filter as well as those unfiltered). The other point is that 3kHz speech almost always has response to 3.5 or even 4.0 kHz at reduced amplitude due to finite line cutoff slope, finite IF skirt selectivity, etc., and that the human ear can make use of this reduced response to a remarkable degree. To utilize the results of the Articulation Index method of evaluating speech channels [7] a band of frequencies must be at least 30 dB down to not influence intelligibility. But for the case of digital systems, there is no possibility of frequencies

---

* In the radio communications situation of the speech signal vs constant per-cycle noise power case, the tradeoff is a little different. In that case, for very marginal S/N conditions, intelligibility is better with lower bandwidth (down to a minimum of about 2.4 kHz) than it would be with greater bandwidth-because widening the bandwidth at upper frequencies would harm intelligibility more due to noise than help it due to greater speech information.

above 1/2 of F_s (sampling frequency) being transmitted except by aliasing which is usually undesirable * so that if the sampling frequency were made the minimum required for 3kHz transmission i.e. 6kHz, there would be no transmission beyond that frequency other than aliasing. This is quite different from normal audio transmission, which typically wouldn't be down 30 dB until at least 3,500 or 4,000 Hz. The importance of this narrow range between 3,000 and 4,000 Hz lies mainly in the noise burst consonants such as the "S" and "SH", "T" and "K" sounds, but the presence of the higher frequencies on voiced sounds does also substantially improve quality, leading to a more "open" sound and less of a telephone-like "restricted" quality. In addition to the improved quality which may do something to offset the loss of quality suffered with LPC, it has also been established that the increased bandwidth improves as measured by the Diagnostic Rhyme Test (DRT) intelligibility, although there may be a point when more than 10 Tap LPC would be required to take full advantage of it. However, the point of this discussion is to dispel the idea that there is nothing worth preserving of the speech spectrum above 3.0 kHz, and to emphasize that the audio input circuitry should be as flat as possible to the upper limit of 4kHz, or preferably somewhat beyond, to avoid any adverse effects associated with the beginning of rolloff. The main determinant of rolloff will thus be the antialiasing filter, which can be carefully controlled as to rate, phase shift, etc. to optimize the system.

---

* Some recent experiments by Kang have shown that careful use may be made of intentional aliasing in certain circumstances.

3.3 Automatic Gain Control Circuits

The objective of Automatic Gain Control circuits is to optimize the voice processor input level to produce the best quality speech possible. The use of 12 bit input A/D conversion in present processors is a great improvement over previous 8 bit digital vocoders and allows the use of a limited AGC range of about 30 dB to optimize the 30 dB dynamic range which occurs within a single speaker plus the additional 20 dB speaker-speaker range to the voice processor. The AGC might not be necessary at all for the above variation except for additional factors which are somewhat special to military applications. The main ones are: (1) The extensive use of noise cancelling microphone handsets on military platforms (first order gradient "confidencer" mikes are routinely used, even in some quiet locations) which are more sensitive to mouth-to-mike spacing than normal handsets and thus produce more working level variation than the normal telephone and (2) The wide variation in background noise which greatly influences speaking volume due to the normal (and proper) human tendency to "speak above the noise". It is estimated that these variables produce as much as 20 dB additional variation, necessitating some sort of AGC action. Unfortunately, most AGC systems produce bad effects which may hurt performance in some cases, while helping it in most other cases. Fast attack, fast release AGC's (such as are used in CB radio "Power Mikes") may blur syllable distinctions, certainly not a desirable effect for LPC systems. Fast attack, slow release AGC (such as typically used for small cassette recorders) are thought to be unapplicable to military systems because of the possibility that heavy gunfire may produce short impulse noise which could keep the gain from ever rising to a usable level for voice. The system specified for developmental ANDVT, which provided that only one 4 dB increment could be made in gain every 1 to 5 seconds was a serviceable compromise, but led to some dissatisfaction due to the long initial adjustment

time. Further improvements made on this system included a faster time increment (currently 32 ms or 1.5 frames) which was made much more acceptable by allowing adjustment to take place only during voiced frames, thus mitigating the problem of "pumping up" during pauses and silence. While no solution is ideal, this evolution seems relatively satisfactory so far.

In summary, it appears that the AGC is a necessary evil whose design is dictated as much by circumstances to avoid, as detailed above, as by the desired reduction in input dynamic variation.

3.4 Sidetone Circuits

Sidetone is defined as the sound of the speaker's own voice heard in the earpiece of a handset during transmission. In the telephone it is normally added to reception from the other end, including line noise, interruption from the other talker, etc.* However, in the telephone, a special circuit is used to reduce sidetone substantially, otherwise, the talker's own voice would be much louder than the voice from the other end due to line attenuation. It has been found that the relative strength of sidetone influences how loud a person talks [8] and this information is used in telephone design to assist in disciplining the user's talking level. However, sidetone may have other effects. For example, the absence of any sidetone is interpreted by most persons as a sign that the telephone is inoperative i.e. dead. More importantly, the quality of sidetone gives the talker some idea of the quality of the connection, and to the extent that the system is reciprocal, i.e. the same in both directions, may be a good guide as to how to talk. For example, a very noisy

---

* This is true only over true full duplex telephone circuits which some modern telephone circuits are not; notably, some satellite circuits and others with automatic echo suppressors.

line, as evidenced by noise in the sidetone, will usually encourage the talker to talk louder; a line with echo may influence the talker to talk more slowly and distinctly, and so forth. There are military voice communications systems, particularly half duplex ones, that provide no sidetone whatsoever, and experienced communicators are accustomed to this situation and use them satisfactorily. One argument in favor of this is that it would be impossible to provide sidetone which represented a true picture of the quality of the transmission link until the reception of the other party provides this information on how noisy, distorted, interference plagued, etc. the particular circuit is (and only then to the extent that it is bi-directionally symmetrical). This would seem to favor eliminating sidetone from half duplex systems completely. The counter argument is based on two points. One is that a wider application of voice systems beyond just experienced communicators is certain with the increased use of secure voice (rather than record) systems. The other argument lies in the unique characteristics of the low bit rate voice coder (LPC Vocoder) system. There seems to be general agreement among those who pioneered these systems and have had the opportunity to observe persons unfamiliar with them try to use them for the first time, that a short familiarization period is needed for a fair percentage of users. This familiarization usually seems to take the form of either reacting to the received synthetic voice in such a way as to attempt to make the outgoing voice sound clearer at the other end by talking more distinctly (and perhaps more slowly or more evenly) or, in some cases, needing to be coached by the person running the experiment as to how to improve his or her speaking style for best transmission. Normally not more than 1 to 10 minutes is required for this accommodation. Also, once learned, it appears to be remembered as long as the talker knows what type of system he is using. This has led to the incorporation of a slight delay (in the order of 30 ms) in the sidetone of service test models of the ANDVT voice processors

as an option. This delay value was chosen to be the same as the echo length that the Bell System permits on telephone lines without requiring echo suppression, and a slight amount of reverberation is also added giving a suggestion of very low level reflections, as would occur in real lines. It is hoped that this would encourage the talker: (1) if he is experienced with this type LPC system, to know that he is talking over such a system (rather than other types of systems that may be accessed from the same subscriber terminal) and thus apply the talking technique most suitable for the system or, (2) if he is not familiar with LPC systems, to be influenced to talk more slowly and clearly as one might do over a line with some echo [9]. The final results and eventual incorporation of such a feature in production equipment await the results of testing and user experience.

In the receive circuit, a small amount of reverberation, which is barely noticeable and has been tested to be sure it does not negatively affect intelligibility, has been incorporated. This reverberation, which amounts to a 30% feedback (-10 dB) of 15 ms delay, was included to compensate for the fact that the synthetic voice lacks the normal acoustic "liveness" associated with speech pickup. This liveness is effectively stripped by the analysis process, which transmits only one pitch impulse per pitch period. Other echoes or reverberations are effectively stripped off, with the only effect transmitted being that of the spectrum change due to cancellation, which is very slight. Another reason for including this reverberation is that it helps the transmit and receive sidetone to have similar characteristics, and hence appears to some observers to lessen the amount of mental accommodation necessary by the talker to the system between transmit and receive modes. As in the case of sidetone delay, final inclusion of the receive reverberation will be based on service test and user tests and observations.

Implementation of the sidetone delay (transmit) and reverberation (transmit and receive) was originally based on Charge Coupled Device (CCD) implementation, with the same circuitry time-shared between transmit and receive, but with delay used in transmit only. Present implementation uses software implementation for reverberation in the receive path only, to eliminate any loss of communication capability in case of CCD chip failure. Future designs for transmit delay would require external digital implementation or a full time D/A converter if done in software. It is recommended that such a design change be delayed until results of field use have been obtained to determine if sufficient motivation is present to overcome any design difficulties associated with incorporation, and both design and production costs.

3.5. Output Circuits

Many of the design criteria for output circuits are the same as for input circuits, and the reader is referred to that section for considerations regarding the choice of IC OP AMPS versus discrete components. One additional criterion relates to the higher peak factor typical of LPC synthetic speech compared to human speech. It has been commonly observed that the speech waveform of LPC speech damps out somewhat faster than normal speech. This may be due to slightly wider formant bandwidth produced, by a less than a fully adequate number of poles for perfect spectrum reproduction, or to the fact that in some systems the excitation function is more instantaneous than in normal speech (particularly normal speech with some room reverberation as commonly observed); but in any case, the problem for the designer is that the output level is usually specified in dBm (RMS power). Thus for synthetic speech, the peak factor will be somewhat higher than the 10 dB used for input speech. It is recommended that a peak factor of 13 dB be used to allow for

this, otherwise there is a danger of clipping the first peak of the formant ringing cycle, with the result of an "edgy" or clipped sound, and a possible loss of intelligibility due to spectral fill-in of pitch harmonics of the clipped signal. For an output signal level of OdBm, which is -2 dB below 1 volt (at 600 ohms), and a peak factor of 13 dB, as recommended, the peak excursion will be as follows:

Peak for sine wave, OdBm across 600 ohms =
1.1 Volts (0.78 VRMS,2.2V Peak-to-Peak (P-P))

Peak for synthetic speech, OdBm =
5.0 Volts (3.6 VRMS, 10.0V Peak-to-Peak (P-P))

It can thus be seen that this maximum voltage cannot be provided by a transistor operated from a single 5V supply (unless a stepup transformer is used); but the danger here is that one might think it could be, based on the following assumptions: (a) The output speech is more or less sinusoidal, so that only about 2.2V P-P might be required, and (b) If the amplifier is designed to take the full D/A converter output without clipping, it may still be impossible to get anything like OdBm for speech signals because of insufficient gain. Therefore, it is recommended that either sufficient supply voltage be used to ensure 10 V P-P capability at 600 ohms load, or that a step-up transformer be used by loading the output circuit with lower impedance so a smaller voltage swing will be sufficient. For example, for a 150 ohm primary winding, 6dB less swing is needed, or only 5 V P-P which may be close to being realizable with a 5 V supply.

In regard to the transformer, the same considerations apply as to the input transformer, and the reader is referred to the section on input coupling. In most cases, the same transformer design should be equally applicable for both input and output circuits.

4.0 Summary and Conclusions

The main points of the report are:

(a) The ear and past experience with "voice circuits" are a poor guide to audio system design for LPC vocoder type digital voice processors.

(b) There are certain pitfalls of miniaturization, particularly as it affects audio input and output transformer design, and unless these pitfalls are avoided, the performance may be seriously impaired by these transformers.

(c) There are also some pitfalls in using IC operational amplifiers for the audio amplifiers. These problems are covered in the report and Ref. 6, which is practically required reading before using IC OP AMPS for critical audio applications.

(d) There are pitfalls in AGC designs as used for voice processors, particularly for military platforms which tend to be more demanding than, say, civilian office-to-office telephone service.

(e) A general discussion of sidetone circuits is given along with a description and brief rationale for some special circuits used in a particular system (ANDVT).

(f) The output circuits are even more critical then input since peak factor of synthetic speech is typically higher than real speech. Thus more capable output amplifiers are needed if typical 0 dBm output level is to be obtained.

REFERENCES

1. Guttman, N. and Flanagan, J. L. "Pitch of High-Pass-Filtered Pulse Trains" JASA, Vol 36, (2) p 757-765, 1964.

2. Pollack, I and Pickett, J. M., "Cocktail Party Effect" JASA Vol 29, p 1262, 1957.

3. Performance Specification for ANDVT Tactical Terminal (TACTERM) #TT-B1-4210-0087, 31 May 1978. Joint Tactical Communications Office, Fort Monmouth, NJ, Sec 3.2.1.2.1.8 "Distortion in Input Circuits" (p.87).

4. Otala, M. "Transient Distortion in Transistorized Audio Power Amplifiers" IEEE Trans. Audio and Electroacoustics, Vol AU-18, pp 234-239 (1970).

5. A Partial discussion is contained in Audio Cyclopedia, 2nd Ed, by *Howard E. Tremain.*

6. Jung, Walter G. Audio IC Op-Amp Applications 2nd Ed. Howard W. Sams and Co. Ind., Indiana, 1981.

7. Kryter, K. D. "Methods For the Calculation and Use of the Articulation Index" Jour Acous Soc Am, 34, 1689-1697 (1962).

8. Richards, D. L. Telecommunications by Speech John Wiley & Sons, NY, Halsted Div., 1973, p 308.

9. Schmidt Nielsen, A. and Coulter, D. C. "Effect of Modest Sidetone Delays in Modifying Talker Rates and Articulation in a Communications Task" Proc. of June, 1979 meeting of ASA, Cambridge Mass.